

Program Report on NASA Subcontract

**Communications and Information Research:
Improved Space Link Performance via
Concatenated Forward Error Correction coding****I. ABSTRACT**

With the development of new advanced instruments for remote sensing applications, sensor data will be generated at a rate that not only requires increased onboard processing and storage capability, but imposes demands on the space to ground communication link and ground data management-communication system.

Data compression and error control codes provide viable means to alleviate these demands. Two types of data compression have been studied by many researchers in the area of information theory: a lossless technique that guarantees full reconstruction of the data, and a lossy technique which generally gives higher data compaction ratio but incurs some distortion in the reconstructed data. To satisfy the many science disciplines which NASA supports, lossless data compression becomes a primary focus for the technology development. Recently, Yeh and Miller [1,2] have shown significant research results in this area using the *Rice algorithm*. The result has been tested for following various applications: (1) Landsat-D Thematic Mapper over Sierra Nevada at 30m ground resolution in band 1 with wavelength region of 0.45 - 0.52 μm ; (2) Soft X-ray Telescope(SXT) image, in the wave length region of 3-60 Angstrom, acquired on SolarA Mission launched in '91; (3) Acousto-Optical Spectrometer(AOS) data, representative of what has been acquired on the Sub-millimeter Wave Astronomy Satellite(SWAS) launched '95. Two traces of 1450 data are in the upper graph and the expanded view in the lower graph; (4) Magnetic Resonance Imaging(MRI) data over the human brain area; (5) Seismic trace acquired in Japan. The compression results show that the extended Rice algorithm is well adapted to various types of sensor data.

On the other hand, while transmitting the data obtained by any lossless data compression, it is very important to use some error-control code. For a long time, convolutional codes have been widely used in satellite telecommunications. To more efficiently transform the data obtained by Rice algorithm, it is required to meet the a posteriori probability (APP) for each decoded bit. A relevant algorithm for this purpose has been proposed by Bahl et al [3]. This algorithm minimizes the bit error probability in the decoding linear block and convolutional codes and meets the APP for each decoded bit. However, recent results on iterative decoding of "Turbo codes", which have achieved low error probabilities at rates well beyond R_0 , turn conventional wisdom on its head and suggest fundamentally new techniques [4].

During the past several months of this research, the following approaches have been developed: (1) a new lossless data compression algorithm, which is much better than the extended Rice algorithm for various types of sensor data, (2) a new approach to determine the generalized Hamming weights of the algebraic-geometric codes defined by

a large class of curves in high-dimensional spaces, (3) some efficient improved geometric Goppa codes for disk memory systems and high-speed mass memory systems, (4) a tree based approach for data compression using dynamic programming. We strongly believe that the research on lossless data compression and error-correcting codes has now reached a stage of very exciting prospects for many commercial, government and defense applications.

II. PROJECT REPORT

II.A. ACCOMPLISHMENTS

1. Personnel

Dr. T.R.N. Rao, Dr. G.L. Feng and G. Seetharaman

Dr. Rao and Dr. Feng investigated the problem of improved space link performance via concatenated forward error correction coding. They derived a new lossless data compression algorithm and developed a new approach to determining a lower bound on the generalized Hamming weights of algebraic-geometric codes defined from a large class of curves in high-dimensional spaces, and worked on developing efficient improved geometric Goppa codes for disk memory systems and high-speed mass memory systems. They provided research support for one post-doctoral and two Master's students.

The graduate students who worked on this project include:

Dr. Xinwen Wu, Ph.D. Mathematics, 1995, post-doctoral.

Mr. Wenji Jin and Mr. Zhiyuan Li, Master students of computer science.

2. Papers (showing acknowledgement of NASA Grant Support)

- [1] G.L. Feng, T.R.N. Rao, and G.A. Berg, "Generalized Bezout's theorem and Its Applications in coding theory" submitted to *IEEE Trans, on Information Theory*.
- [2] G.L. Feng, X.W. Wu, and T.R.N. Rao, "New Double-Byte Error-Correcting Codes for Memory systems" will be submitted to *IEEE Trans, on Information Theory* and 1997 IEEE International Symposium on Information Theory.
- [3] X.F. Shi, X.W. Wu, G.L. Feng, and T.R.N. Rao, "The Applications of Generalized Bezout's Theorem to the Codes from the Curves in High Dimensional spaces," will be submitted to *IEEE Trans, on Information Theory*.
- [4] X.W. Wu, G.L. Feng, and T.R.N. Rao, "The Weight Hierarchy and Chain Condition of a Class of Codes from Varieties over Finite Fields," invited talk at *Thirty-Fourth Annual-Allerton Conference on Communication, Control, and Computing*, Oct. 3-7, 1996.
- [5] G.L. Feng, T.R.N. Rao, and E.W. Hinds "Rice-Like Data Compression for Remote-Sensing and Other Applications", preparing.
- [6] G.L. Feng, T.R.N. Rao, and E.W. Hinds "On the Optimization of Rice-Like Algorithm for Spacecraft Television Data," preparing.

- [7] G. Seetharaman and B. B. Nair, "Requantization of digital Images for Data Compression: A Dynamic Programming Solution," submitted to *SIAM symposium on Discrete algorithms*,
- [8] G. Shivaram and G. Seetharaman, "Data Compression of Discrete Sequence: A Tree Based Approach Using Dynamic Programming,"

3. Other Research Support

- (1) G.L. Feng and T.R.N. Rao, "Generalized Hamming Weights of Algebraic-Geometric Codes and Asymptotically Good Algebraic-Geometric Codes," The National Science Foundation, \$260,000.00, September 1995 - August 1998.
- (2) G.L. Feng and T.R.N. Rao, "Improved Algebraic-Geometric Codes," LEQSF, \$63,000.00, June 1994 - August 1996.

4. Specific Technical Accomplishments

The overall goal of this project is the development of some more efficient data compression algorithms for remote-sensing and other applications, and the construction of improved generalized Goppa codes. An outline of the approach towards solving this problem is presented in the following discussion.

4.1 New Algorithms for Data Compression

Here, we give some new algorithms, which are different from the known algorithms. The simulation results show that our new algorithms perform better than the Rice algorithm.

The Concept

Let $\tilde{\delta}^n = \delta_1 * \delta_2 * \dots * \delta_J$ be the data to be compressed. If we consider each δ_i as a column vector, $\tilde{\delta}^n$ can be considered as a matrix. That is

$$\tilde{\delta}^n = \begin{matrix} & \delta_1 & \delta_2 & \dots & \delta_J \\ \begin{matrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{matrix} & \boxed{\text{binary matrix}} & \begin{matrix} \uparrow \\ \log J = n = 10, \\ \downarrow \end{matrix} \end{matrix}$$

$J = 1024$

where l_i is the i -th row of the matrix.

The coding problem can be regarded as the following problem. Let d_1, d_2, \dots, d_n be the symbols corresponding to the 1's at each row and ',' be the symbol for the column change. Then, the binary matrix can be reduced to a sequence

on the set of $\{d_1, d_2, \dots, d_n, ', '\}$. For example, consider binary matrix

$$\begin{array}{cccccccccccccccc}
 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & \longrightarrow & d_1 \\
 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & \longrightarrow & d_2 \\
 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & \longrightarrow & d_3 \\
 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & \longrightarrow & d_4 \\
 & & & & & & & & & & & & & & & & \longrightarrow & ,
 \end{array}$$

Let $\Sigma = \{d_1, d_2, \dots, d_n, ', '\}$. The above binary matrix is reduced to the following sequence

$$d_1 d_4, d_1 d_2 d_4, d_1 d_2 d_3, d_3 d_4, d_1 d_4, d_1 d_2 d_4, d_1 d_2 d_4, d_4, \dots$$

Our new algorithm based on the above idea can be described as follows.

Algorithm 1

Step 1: Calculate the weight (the number of 1's) of each row: w_1, w_2, \dots, w_n .

Step 2: Sort the weights such that $w_{i_1} \geq w_{i_2} \geq \dots \geq w_{i_n}$. Let $w_{i_0} = J$. Calculate $W = \sum_{p=0}^n w_{i_p}$.

Step 3: According to w_{i_p}/W , we find a variable-length coding scheme for encoding the 1's in each row and the column change symbol ', ' by using Huffman code.

Step 4: Encoding the columns of the binary matrix one by one, from left to right. The final code is a sequence of Huffman code words.

Remark 1: In Step 3, Huffman code is not necessarily the only coding scheme for encoding the symbols of 1's in each row and the column change. Many other codes, such as Algebraic Geometric codes, can be used here.

If we investigate the binary matrix carefully, we find that the above algorithm can be improved in two different ways.

Improved Algorithm A

The first three steps are the same as those in Algorithm 1.

Step 4: Let $l_0^{(k)}, l_{n-k}^{(k)}, \dots, l_1^{(k)}$ be the lengths of the Huffman code words for $w_{i_0}, w_{i_{n-k}}, \dots, w_{i_1}$. Let k^* be the index such that

$$F(k^*) = \min\{F(k) | k = 0, 1, \dots, n-1\},$$

$$\text{where } F(k) = J l_0^{(k)} + \sum_{p=1}^{n-k} l_p^{(k)} w_{i_p} + kJ.$$

Step 5: Read the $n - k^*$ rows which correspond to $w_{i_{n-k}}, \dots, w_{i_1}$ as the most significant $n - k^*$ bit samples coding scheme. The other k^* rows are read as the k^* least bit samples.

Improved Algorithm B

The first three steps are the same as those in Algorithm 1.

Step 4: Exchange the row such that $w_1 \leq w_2 \leq \dots \leq w_n$. Let d_i present the 1's in the i -th row,

Step 5: (a) If the last 1 of the current column is in the i -th row, i.e. $d_i = 1, d_j = 0, j > i$, and the first 1 of the next column is in the k -th row, i.e. $d_k = 1, d_j = 0, j < k$. Thus, the encoding sequence is $\cdots d_i, d_k \cdots$. If $i \leq k$, then the ',' between d_i and d_k can be removed.
(b) If for the current column, the sequence is $\cdots \underbrace{d_i d_{i+1} \cdots d_{i+p}}_{d_i, d_{i+p}} d_k$, where $j > i + 1, i + p > k + 1$. Then, $d_i d_{i+1} \cdots d_{i+p}$ can be represented by d_{i+p}, d_i .

These algorithms have been implemented by using simulation. The results show that the new algorithms are better than the extended Rice algorithm. More details will be provided in the next technical report.

4.2 Generalized Hamming Weights of AG codes

For error-correcting codes, the minimum distance is one of most important parameters. It is used to measure the code's capacity of correcting errors or detecting errors or both. The minimum distance d of a linear code C is defined by

$$d = \min_{\substack{\mathbf{u}, \mathbf{v} \in C \\ \mathbf{u} \neq \mathbf{v}}} \{d(\mathbf{u}, \mathbf{v})\},$$

where $d(\mathbf{u}, \mathbf{v})$ expresses the Hamming distance between \mathbf{u} and \mathbf{v} .

For an $[n, k]$ linear code, we can consider its generalized Hamming weights, which are the generalization of minimum distance.

Both the determination of the minimum distances and the determination of weight hierarchy for linear codes in full are difficult. A more modest goal is to find acceptable bounds on these weights. The weights of geometric Goppa codes were discussed in several papers. The bounds on the minimum distance and the generalized Hamming weights of the codes defined on the curves in two-dimensional space were given by Feng, Rao and Berg in their paper.

We consider the codes defined on the curves in n -dimensional space. We are now interested in the following irreducible space curves:

$$\begin{cases} f(x_1, x_2) = 0, \\ f(x_1, x_2, x_3) = 0, \\ \cdots \cdots \cdots \cdots \\ f_{n-1}(x_1, x_2, \cdots, x_n) = 0, \end{cases}$$

where

$$f_s((x_1, x_2, \cdots, x_{s+1}) = x_s^{a_s} + x_{s+1}^{b_s} + g_s(x_1, x_2, \cdots, x_{s+1}),$$

$$\gcd(a_s, b_s) = 1 \text{ and } \deg g_s(x_1, x_2, \cdots, x_{s+1}) < \min\{a_s, b_s\}.$$

Let a point $p = (i_1, i_2, \cdots, i_n)$ in R^n represent a monomial $x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$. We define the weight of the monomial as follows:

Definition 1 For a n -dimensional monomial $x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$, we define its weight as

$$w(x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}) = \sum_{j=1}^n \left(\prod_{k=1}^{n-j} b_k \prod_{k=n-j+1}^{n-1} a_k \right) i_j.$$

Lemma 1 $D_{\{[x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}]\}} \leq \sum_{j=1}^n 4^{n-j} 5^{j-1} i_j.$

Theorem 1 $D_p^{(r)} \leq w(\mathbf{h}_r) - w(\mathbf{h}_p).$

Theorem 2 Let C_r be a $[4^{n+1}, 4^{n+1} - r]$ code defined by parity-check matrix $\mathbf{H}_r = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_r]^T$. Then

$$d_h(C_r) = h + r, \text{ if } h \geq w(\mathbf{h}_r) - r + 2.$$

This work has resulted in a paper titled “The Applications of Generalized Bezout’s Theorem to the Codes from the Curves in High Dimensional spaces,” by X.F. Shi, X.W. Wu, G.L. Feng, and T.R.N. Rao, which will be submitted to the *IEEE Transactions on Information Theory*.

4.3 Chain Condition of a Class of Codes from Varieties

The generalized Hamming weights and weight hierarchies of linear codes were first introduced by Wei, which are fundamental parameters related to the minimal overlap structures of the subcodes and very useful in several fields. It was found that the chain condition of a linear code is convenient in studying the generalized Hamming weights of the product codes. We considered a class of codes defined over some varieties in projective spaces over finite fields, whose generalized Hamming weights can be determined by studying the orbits of subspaces of the projective spaces under the actions of classical groups over finite fields, i.e., the symplectic groups, the unitary groups and orthogonal groups. We gave the weight hierarchies of the codes from Hermitian varieties and proved that the codes satisfy the chain condition.

Consider the finite field \mathbb{F}_{q^2} with q^2 elements, where q is a power of prime. \mathbb{F}_{q^2} has an involute automorphism

$$a \longmapsto \bar{a} = a^q.$$

The fixed field of this automorphism is \mathbb{F}_q .

Let $k = \nu + l$, where $\nu > 0$, $l \geq 0$, and

$$I_{(\nu, l)} = \begin{pmatrix} I^{(\nu)} & \\ & 0^{(l)} \end{pmatrix}.$$

The set of points ${}^t(x_1, x_2, \cdots, x_k)$ satisfying

$$(x_1, \cdots, x_k) I_{(\nu, l)} \overline{{}^t(x_1, \cdots, x_k)} = 0$$

is a Hermitian variety in $PG(k-1, \mathbb{F}_{q^2})$, when $l = 0$, it is a nondegenerate Hermitian variety, and when $l > 0$, it is a degenerate Hermitian variety.

We denote this Hermitian variety by $I_{(\nu,l)}$. Let $n = |I_{(\nu,l)}|$ be the number of points lying on $I_{(\nu,l)}$ in $PG(k-1, \mathbb{F}_{q^2})$. For each point of $I_{(\nu,l)}$, choose a system of coordinates and regard it as a k -dimensional column vector. Arrange these n column vectors in any order into $k \times n$ matrix, denote it also by $I_{(\nu,l)}$. It can be proved that $I_{(\nu,l)}$ is of rank k . Hence $I_{(\nu,l)}$ can be regarded as a generator matrix of a q^2 -ary projective $[n, k]$ -code, which will be denoted by $C_{(\nu,l)}$.

By studying the orbits of subspaces of the projective spaces under the actions of unitary groups over finite fields, we determined the complete weight hierarchy of $C_{(\nu,l)}$, and proved that $C_{(\nu,l)}$ and $C_{(\nu,l)}^\perp$ satisfy the chain condition. As a corollary, we showed that when ν is even and $r \leq \nu$, $d_r(C_{(\nu,l)})$ meets the Griesmer-Wei bound.

Many applications of generalized Hamming weights are known. They are useful in cryptography, in trellis coding, and in truncating a linear block code, etc.

Our results opens the possibility of determining the complete weight hierarchies of any product code by a code from Hermitian variety or its dual code and other linear code.

This work resulted in a paper titled "The weight hierarchies and chain condition of a class of codes from varieties over finite fields" by Xinwen Wu, G. L. Feng and T. R. N. Rao. This paper will be presented at *Thirty-Fourth Annual-Allerton Conference on Communication, Control, and Computing*, Oct. 3-7, 1996.

4.4 Efficient Error-Correcting Codes for Memory Systems

Error-correcting or error-detecting codes are useful in computer semiconductor memory subsystems, which can be used to increase reliability, reduce service costs, and maintain data integrity. It is well known that the single-byte error-correcting and double-byte error-detecting (SbEC-DbED) codes have been successfully used in computer memory subsystems. For a linear block code over the finite field $GF(q)$ of q elements, where q is a prime power, if its minimum distance is equal to or greater than d , then the code is capable of correcting $\lfloor \frac{d-1}{2} \rfloor$ byte errors and detecting $\lfloor \frac{d}{2} \rfloor$ byte errors. Thus the minimum distances of linear codes which are capable of correcting single byte errors and detecting double byte errors are equal to or greater than four, and the minimum distances of the codes which can correct double byte errors are equal to or greater than five.

Let $U_q^m(I)$ be a cyclic code over $F = GF(q)$, where $q = 2^i$, with a string $I = \{1, \frac{(q^m+q)}{2}\}$, and $U = U_q^m(I, F^{m-1})$ be the corresponding punctured code with length $n = q^{m-1}$ defined on a $(m-1)$ -dimensional subspace F^{m-1} of F^m . A class of codes over $GF(2^i)$ with minimum distance ≥ 5 was constructed by adding some parity checks to U . And when q is odd, a class of codes with minimum distance ≥ 5 was also constructed by a similar method. The above codes were constructed by Dumer. According to Dumer, if q is even, when $n = q^2$, $r \leq 7$, when $n = q^3$, then $r \leq 9$; and if q is odd,

when $n = q^2$, $r \leq 7$, when $n = q^3$, then $r \leq 8 \dots$. Dumer's codes are known to be optimal in the sense that no other double-byte error-correcting codes with the same code lengths have fewer number of parity checks. But unfortunately, the codes were defined only over $GF(q)$, when q is odd.

It is well known that in the computer systems the codes over $GF(q)$ with $q = 2^i$ are useful. In this research, we construct a class of double-byte error-correcting codes over $GF(2^i)$, which have the same parameters of Dumer's codes over $GF(q)$ when q is odd. And we also obtain a decoding procedure of our codes.

Dumer's codes have the parameters:

When q is odd,

$$n = q^{m-1}, \quad r \leq 2m + \lceil \frac{m-1}{3} \rceil, \quad d \geq 5, \quad m = 2, 3, \dots$$

When q is even,

$$n = q^m, \quad r \leq 2m + \lceil \frac{m}{3} \rceil + 1, \quad d \geq 5, \quad m = 2, 3, \dots$$

Our main result is: Over finite field $GF(q)$, q is odd or even, we constructed linear codes with the parameters:

$$n = q^m, \quad r = 2m + \lceil \frac{m}{3} \rceil + 1, \quad \text{and } d \geq 5, \quad m = 3, 4, \dots$$

The development of this research opens the possibility of raising the speed in communication systems and computer memory systems.

This work resulted in a paper titled "New double-byte error-correcting codes for memory systems" by G. L. Feng, Xinwen Wu and T. R. N. Rao. This paper will be submitted to *IEEE Trans, on Information Theory* and 1997 IEEE International Symposium on Information Theory.

4.5 Discrete Algorithms for Data Compression

Our research in data compression described in this section follows a discrete algorithmic approach. Our goal is to find the intrinsic measure of compressibility of a given data set. To accomplish this, we have developed a suitable representation of the problem as an integer-programming, resource-allocation problem.

Our approach is very different from the popular approaches such as Wavelet Transform based Coding, Cosine Transform coding etc., all of which make certain implicit assumptions about the spatial characteristics of the data. Though such simplifications result in attractive performance over a class of images specific to each method, there is no concrete method available at present for comparing the performance of two such approaches. The basic results from rate distortion theory are often used for this purpose, which does not facilitate conclusive comparison of the merits of two competing methods. We have addressed this problem by designing a method which will

find a truly optimal compression of a given image, satisfy given error bound, and a channel capacity.

Two algorithms have been developed by recognizing compression as a discrete optimization problem, and approaching it with dynamic programming methods. The results have been submitted for publication at the SIAM ACM Symposium on Discrete Optimization Algorithms, Jan. 1997. We are currently applying these methods, and conventional methods, on a large set of images, to establish an experimental test bed for data compression algorithms. Two journal articles are being prepared at present. Our solutions are briefly described below.

4.5.1 Requantization (Adaptive Thresholding) of Digital Data for Data Compression.

We present a simple formulation of image data compression as a discrete optimization problem. In particular, it is proposed to requantize the image using an integer / dynamic programming approach. The goal is to reduce the number of bits required to store or transmit the given image to a remote location. And, it is required keep the net error between the original image and the decompressed version at its minimum for a given channel capacity (integer valued resource). This algorithm is also quite useful for segmenting an image based on gray scale properties, using multiple thresholds.

For example, given a image of $N \times N$ pixels at 8 bits/pixel. A coarse requantization of 1 bit amounts to N^2 bits for the whole image. Our task is to assign, distinct gray-levels, or code words to each of 2^8 input levels, such that the errors due to requantization are minimized. This is an instance of resource allocation problems, with integer valued resources C (or channel capacity), and specific weights (error). We have modeled this problem as a NP-complete problem of finding an optimal cut set of an interval tree, whose top node is the entire range $[0..255]$ of gray-levels, and exactly C/N^2 , and the leaves decimate the entire range into specific number of code-words.

4.5.2 Optimal Partitioning of a Sequence of Numbers for Compression

Our second approach assumes the input data of N numbers as a discrete interval $[0..N]$ and decimates the interval into several non overlapping intervals. Each interval is approximated by one or two metrics, for example, the average value, moments etc., from which the members of each interval can be reconstructed. The total number of subsequences has a direct connection with channel capacity, and the error that results due to approximating each interval by the average, second-moment, etc. corresponds to the weight. Then our compression strategy is to find an optimal partitioning of the input sequence for a given channel capacity. This problem is also mapped into finding an optimal cutset of an interval tree. Experimental implementation has been completed. A journal article is under preparation for publication in the *IEEE Trans. on Image Processing*.

II.B. NEXT REPORT PERIOD

1. Personnel (USL Subcontract)

Dr. Rao and Dr. Feng, as well as their post-doctoral Dr. Wu, will continue to work on this project. Their graduate students, Mr. Shi, Mr. Jin and Mr. Li will still be involved in the project.

2. Specific Technical Accomplishments

In this proposal, we intend to investigate the following problems:

- (1) A new Rice-like data compression algorithm for remote-sensing and other applications has been developed. Some simulations have shown that the results were much better than that by the extended Rice algorithm for some cases. In the next period, we would like to test the performance of the new algorithm on various test imagery. On the other hand, we also would like to improve the Rice-like data compression algorithm by developing a modified Huffman code, which has minimal expected average length and is quasi-uniquely decodable. Quasi-uniquely decodable means that any subset of the ordered codeword sequence can be uniquely decodable. For example, let $abcdefghijk$ be a ordered codeword sequence and $a, b, c, d, e, f, g, h, i, j, k$ are all codewords. $bdghk$ is a subset of this ordered codeword sequence. Obviously, the condition of prefix is stronger than the condition of quasi-uniquely decodable. Thus, the Rice-like algorithm can be improved by using the modified Huffman code. To analyse and test the performance of the improved Rice-like algorithm is also proposed.
- (2) Let $\{a, b, c, d, e, f, g, h\}$ be a source symbol set. Let $cdehbgdefgfahach$ be an original character sequence. Its entropy is 2.952820. If we decode a as 000, b as 001, and so on. The symbol sequence is encoded in a binary sequence. We add (modular two) a known binary sequence to the sequence. Then we get a new binary sequence. Decoding the new binary sequence, we get a new symbol sequence $dhddhgfhdfgdghdd$, whose entropy is 1.849602. We are very interested in this preliminary result, because there may be a new lossless data compression algorithm. We propose to investigate a new lossless data compression algorithm by changing the original binary sequence. This technique can be used in combining with Huffman coding or arithmetic coding such that a new efficient universal lossless data compression can be found.
- (3) We propose to develop a class of fixed-byte error protection codes, that are suitable for the data obtained by the extended Rice algorithm. The data is divided into two segments. The data in one segment is the J sample sequence of the most significant $n - k$ bits samples extracted from the original data. For the data in this part, it is required to be error free. The data in the other segment is the corresponding sequence of the k least significant bit samples of the original data. For the data in this segment, a small error rate is tolerated.
- (4) We propose to investigate a new encoding and decoding scheme that should greatly enhance the likelihood of detecting any single or multiple

bit errors that may occur during transmission and reception of information. The proposed scheme must attain a bit error rate of the order of 10^{-15} to 10^{-17} , and have a minimal implementation overhead.

References

- [1] P. S. Yeh and W. H. Miller, "A Real Time Lossless Data Compression Technology for Remote-Sensing and Other Applications," IGRSS, 94.
- [2] P. S. Yeh, R. F. Rice, and W. H. Miller, "On the Optimality of a Universal Noiseless Codes," The Proceeding of the AIAA Computing in Aerospace 9 Conference, 1993.
- [3] L. R. Bahl, J. Cocke, F. Jeinek, and J. Raviv, "Optimal Decoding of Linear codes for Minimizing Symbol Error Rate," *IEEE Trans. Inform. Theory*, Vol. IT-20, pp. 248-287, March, 1974.
- [4] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding: Turbo codes," *Proc. 1993 IEEE Int. Conf. Commun.*, pp. 1064-1070, May 1993.